Published in Solar Physics, vol 168, n°2, pp. 423-433, 1996

COMPARISON OF NEURAL NETWORK AND McNISH AND LINCOLN METHODS FOR THE PREDICTION OF THE SMOOTHED SUNSPOT INDEX

Françoise FESSANT *France Telecom/CNET, Departement LAB/RIO/TNT, 22307 Lannion Cedex, France*

Catherine PIERRET *CNES, Department CT/TI/MS/MO, 18 Avenue Edouard Belin, 31055 Toulouse Cedex, France; and Observatoire de Meudon, 92195 Meudon Cedex, France*

Pierre LANTOS *Observatoire de Meudon, 92195 Meudon Cedex, France*

Abstract.

In this paper we propose a comparison between two methods for the problem of long-term prediction of the smoothed sunspot index, These two methods are first the classical method of McNish and Lincoln (as improved by Stewart and Ostrow), and second a neural network method. The results of these two methods are compared in two periods, during the ascending and the declining phases of the current cycle 22 (1986-1996). The predictions with neural networks are much better than with the McNish and Lincoln method for the atypical ascending phase of cycle 22. During the second period the predictions are very similar, and in agreement with observations, when the McNish and Lincoln method is based on the data of declining phases of the cycles.

1. Introduction

The prediction of solar activity is of importance for many applications. Space technology needs it to predict satellite orbit, lifetime and shielding. Space manned flights in high-inclination orbits or in trajectory outside the protection of the Earth's magnetic field need monitoring as well as prediction of solar activity. Radio communications, electric power and cable network operations and pipeline maintenance could also be disturbed by solar activity (Lanzerotti, 1983). The predictions are sometimes source of important scientific progress, such as the work done by Ohl (1966, 1976), which has been at the origin of the now widely accepted concept of the extended solar cycle (see, for example, Wilson, 1994).

Long-term predictions (months or years ahead) are frequently applied to the smoothed sunspot index (RI_{12}), in fact a 13-month running mean of RI, or to the formally equivalent centimeter flux at λ = 10.7 cm. A number of methods have been developed (see Denkmayr, 1994, for a comprehensive analysis). Cycle prediction covers two different parts: the prediction of the future evolution ofthe cycle in progress, and the prediction of the following cycles. Both are dealing with different methods, and we consider here only the prediction of the current cycle. We describe first the method proposed by McNish and Lincoln (1949) and improvements of it proposed later. This regression method is considered as one of the best (Hildner and Greer, 1990), and it is in operational use in forecasting centers. We then describe a method developed at Lannion French National Telecommunication Research Laboratory (CNET) using neural

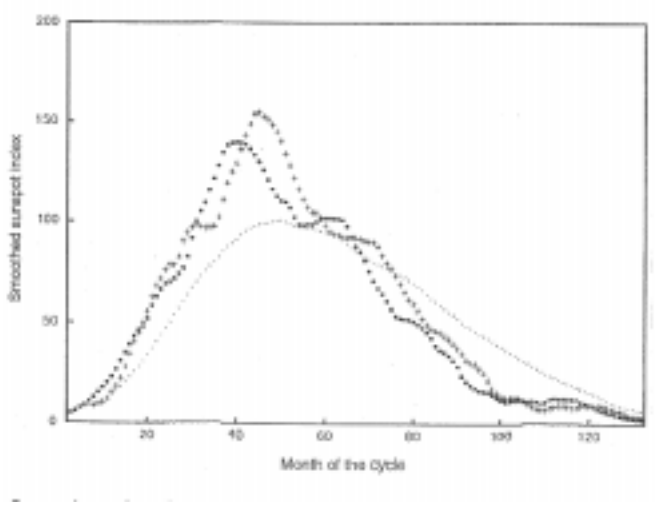

Figure 1 : Comparison of predictions with the McNish and Lincoln method (+) and observations (●.) for cycle 11. The mean cycle is also indicated with a dotted curve

networks (Fessant, 1995). Neural networks are successfully used in various fields and are well suited to resolve problems of solar-terrestrial predictions (Gorney, Koons, and Waltersheid, 1993; Lundstedt, 1993), for which many examples are available but not the equations that govern the system. Recently neural networks have been used for solar cycle prediction by Koons and Gorney (1990), McPherson (1993), and Tian (1996). Both methods have been implemented, and comparison is made of their results for the rising part and for the declining part of cycle 22, which started in September 1986.

2. The Improved Method of McNish and Lincoln

A formula for predicting the smoothed annual sunspot numbers was published by McNish and Lincoln in 1949. A first approximation to the prediction of a future value in a cycle is the mean of all past values for that part of the cycle. This estimate can be improved by adding to the mean a correction factor proportional to the departure of earlier values to the cycle from the mean cycle. The correction factors are determined by the method of least squares. Sunspot data for 1834 though 1943 (cycles 8 to 17) were used, and for statistical reasons data of cycles 1 to 7 are not used for this analysis. After tests, McNish and Lincoln decided to employ only the mean value, corrected by the departure of the mean sunspot index of the preceding year for prediction of future smoothed annual sunspot indices.

This method has been greatly improved by Stewart and Ostrow in 1970. Indeed, they have described an adaptation of the McNish-Lincoln technique to sunspot data spaced at monthly intervals, which allows the prediction of monthly mean values. In their procedure, the effect of short-term fluctuations in the monthly mean values is minimized by the use of 12-month running means of the index. It is necessary with this method to use an extra six months extrapolation when predictions of this smoothed index are being made.

The hypotheses used in the Stewart and Ostrow method are:
(a) The independence of individual cycles.
(b) All the cycles observed since 1749 (the minimum of cycle 1 is in March

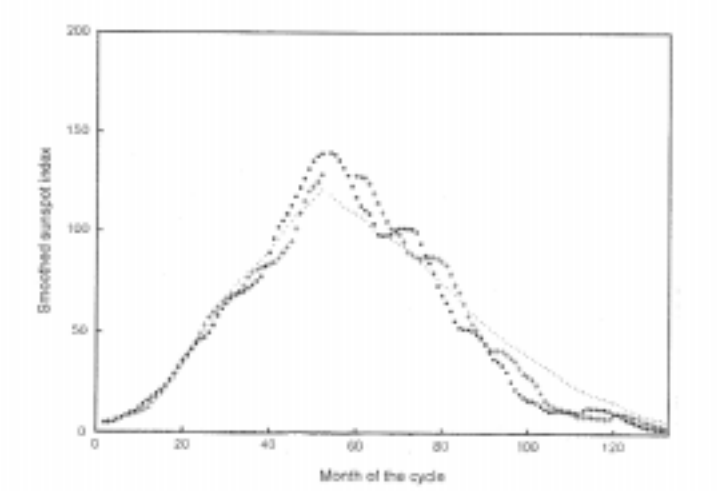

Figure 2 : Comparison of observations (●) with predictions of McNish and Lincoln method when the cycle is separed into two phases (+). The dotted curve is the mean cycle in this case.

1755) belong to the same statistical population. This is the first important difference from the McNish - Lincoln method.
(c) The step size used for the prediction method is one month. This is the second difference with the McNish-Lincoln method, in which the predictions are made one year in advance.

The method implemented follows the approach of Holland and Vaughan (1984)
for the computation of the mean cycle. The steps of the implementation are:
- the mean period (P months) of cycles I to 21 (or 8 to 20 because the result is similar) is calculated;
- each cycle is resampled at P + 1 points (at the same phase of each cycle), by means of an interpolation technique; thus, each cycle has the same length.

Note that by attributing the mean period to the current cycle we infer the minimum date of October 1997 for cycle 22. A mean cycle is computed based on the resampled values. The prediction of a future value in the current cycle k ($x_{j,k.}$ where j is the month) is the mean of all past values for that part of the cycle $X_{j,k-l}$, (the mean cycle value in the month j) added to a correction proportional to the difference between the earlier value of the same cycle k ($x_{j-7,k}$ on account of the ~ix month's delay) and its respective mean $x_{j-7,k-l}$:
a_1 ($x_{j-7,k}$ - $X_{j-7,k-1}$, where a_1 is obtained by minimizing χ^2.

To test the performance of this method, the scheme has been applied to a prediction over all the cycles (1 to 21); we calculate for each cycle the standard jeviation for the predicted values against the observed, giving a standard deviation) of 2.3 (mean is 7.56). This is in satisfactory agreement with the observations. Figure 1 shows the predicted results of cycle 11 compared with the observations and the mean cycle. Another change, already used in the Boulder Warning Center :Greer, 1993), has been implemented, which still improved the preceding one. Sunspot cycles have an average length of about 11 years, but their duration can vary from 8 to 13 years. Their amplitude can also be very different. So, on account of this , variability, the idea is to cut the cycles at the time of maximum. We have now two half-cycles (an ascending phase and a descending phase) for each cycle. The method of McNish and Lincoln is applied to each half-

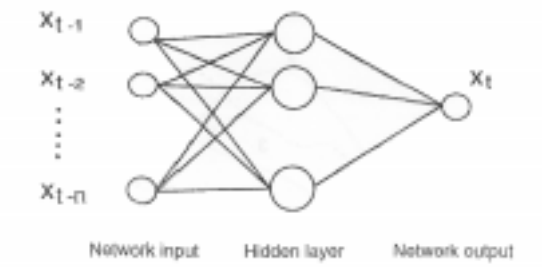

Figure 3 : A simple neural network with one hidden layer of units where all units in one layer are connected to all units in the next layer

cycle. A mean ascending phase is computed based on the resampled values according to the mean period P1 (the average ascending time from minimum to maximum is approximately 50 months) and a mean descending phase is computed based on the resampled values according : to the mean period P2 (the average descending time from maximum to minimum is about 80 months). The inferred minimum date for the current cycle in this case is March 1996. When we separate the ascent and the descent, the predictions over all the cycles are better, especially for the declining phases of the cycles. The standard leviation of the sum of all the half-cycle standard deviations for the predicted values against the observed is 2.5 for the ascending phases (mean is 7.41) and 1.9 for the descending phases (mean is 5.52). Figure 2 shows the predicted results of cycle 11 compared with the observations and the mean descending cycle by using his 'cut at maximum' method.

Nevertheless the method of McNish and Lincoln presents some difficulties. It requires that the month of the last solar minimum be known. Because it uses a number of months after a minimum to ensure that the minimum identified is the real one and not just a subsidiary phenomenon, it is obvious that there will be a significant delay before predictions can commence at the beginning of a cycle. This delay will have to be added to the time-lag created by the forward prediction period and to the extra six months involved. Another difficulty may arise at the end of the cycle if the current cycle is longer than the reference average cycle. If there are no reference data for long cycles, no predictions are possible. Thus, unless modifications to the standard procedure are introduced, predictions cannot be produced in this case for the months around the minimum epoch.

3. Neural Network for Prediction

The prediction problem consists, given the first n values of a one variable time series $\{x_1, ..., x_n\}$, in finding the future values $\{x_{n+1}, x_{n+2}, .. .\}$, where x_t is the series value sample at time t. It has been shown by Takens (1981) that if the series is deterministic, there exists an integer d (called the embedding dimension), an integer δ (a delay) and a function f such that for every $t > d\delta$

$$X_t = f(W_{t-\delta}, x_{t-2\delta}, ... , x_{t-d\delta}) .$$

Neural networks, which can be used as universal function approximators, are used to approximate the function f.

Neural networks consist of a large number of highly connected, nonlinear, simple units. In the models used in predictions we can distinguish 3 types of units:

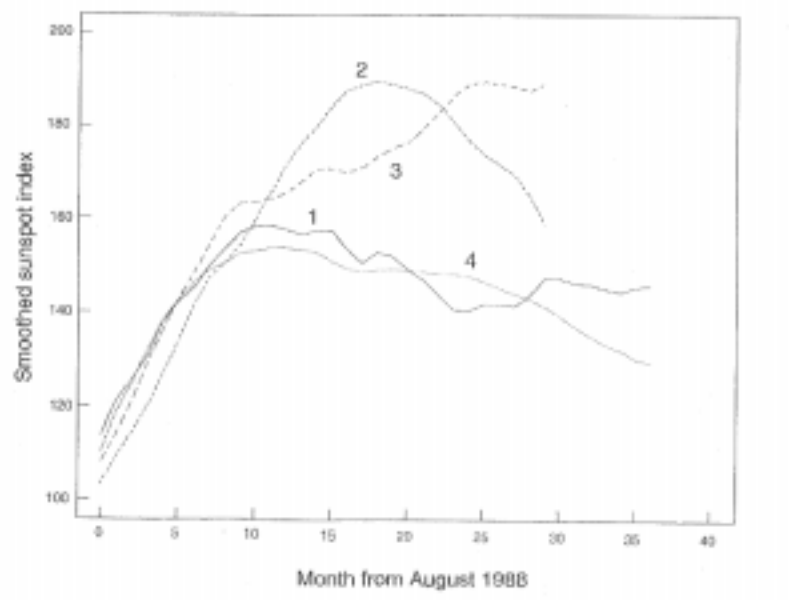

Figure 4 : Comparison of observations for the ascending phase of cycle 22 (curve I), with predictions obtained with the McNish and Lincoln method: using the entire cycle (curve 2) or the ascending phase of cycles (curve 3) and with neural network (curve 4).

- Input units which are set to the previous value of the time series $x_{t-\delta}$, $x_{t-2\delta}$, . . , $x_{t-d\delta}$, where d is the embedding dimension.
- Output units which give the results of the neural network. In the simplest case, we have only one output unit which should return x_t.
- Finally, hidden units which are neither input nor output units, but are used to keep an internal representation of the problem.

Each connection between two units is directed and is given a weight. In fact, the knowledge of the network is kept in these weights. Each hidden and output unit computes its value as the weighted sum of its inputs, passed through a nonlinear function such as hyperbolic tangent.
The idea then is to find, for a given network architecture and a given time series, the weights that minimize a cost which is a function of the difference between the resulting values of the network and the desired values. The time series is generally divided into two parts: a training set and a test set. The training set (for instance the first values of the time series) is used to find the weights by minimizing a cost function, whereas the test set (for instance the last values of the time series) is used to verify the real prediction performance of the network. The most widely used cost function is the least mean-square criterion and the minimization of this cost function is usually done by an iterative procedure which consists of the following steps: first initialize the network weights randomly, then for each example in the raining set, compute the network output while feeding the example as input, compare the resulting output to the target, and apply a correction to all weights vhich minimize the error. One iteration is the presentation of all examples. The procedure could last many iterations. The most widely known learning mechanism or neural networks is the back -propagation rule (Rumelhart, Hinton, and Williams,1986). It is a simple gradient descent technique, which minimizes the cost function in the weight space.

One of the most important features of learning systems is their ability to generalize to new situations. As we have seen, a learning machine such as a neural network is usually trained to

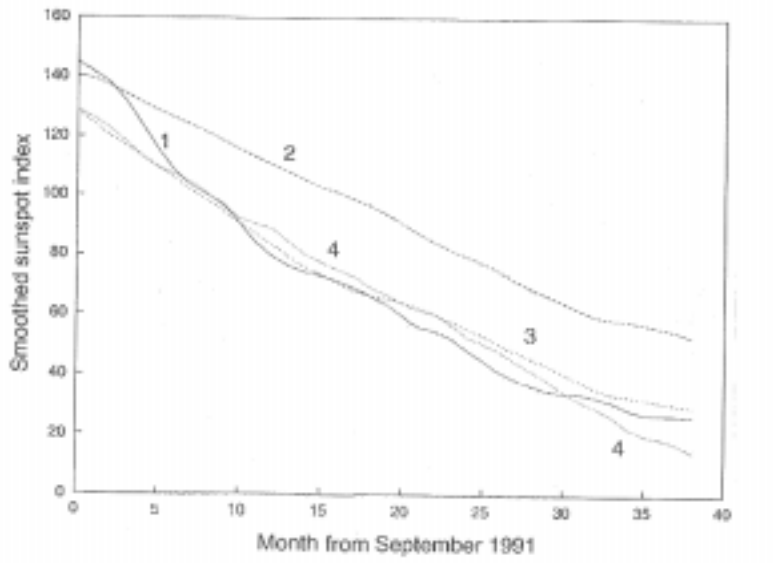

Figure 5 : Comparison of observations for the declining phase of cycle 22 (curve I) with predictions btained with the McNish and Lincoln method: using the entire cycles (curve 2) or the decreasing hase of cycles (curve 3) and with neural network (curve 4).

minimize a cost function over a finite set of examples (the training set), but what we are really interested in is to find a function which minimizes our cost function over all the input domain. Particulary, in prediction problems, we train the network with past examples (thus we minimize a training error) but we really want our network to perform well on future examples (thus, have a minimal generalization error. We usually use a test set (data not used to minimize the cost) to estimate the generalisation error. Theoretical results such as Vapnik (1982) show that the smallest generalization error we can reach is a function of the training set size, the network capacity (which is roughly a measure of the number of free parameters), and the training error. This means that we have to find the best network architecture for a given problem and a given training set size. Many heuristics exist (Fahlman and Lebiere, 1990; LeCun et al., 1990) but it is still a hard problem.

The parameters of the neural network model, i.e., embedding dimension d (or input size of the network), as well as the hidden layer size are usually set by cross validation techniques. The training set (for instance the first values of the time series) is cut into two parts: a new training set and a validation set. Different network models (and more specifically different input size networks) are trained with this new training set; an error on the validation set is computed for each model and these validation errors are compared. We keep the best model (i.e., the model which has the lowest error in the validation set). The input dimension of this model is then a good estimate of the embedding dimension d. For more information about neural networks, see for instance (Hertz, Krogh, and Palmer, 1991).

In this application, we want to predict a future value of the RI12 index in a long range horizon. To solve this problem, we use iterated prediction. This consists in feeding the previously predicted values X_t, X_{t+l}, X_{t+l-1} as inputs of the network to predict X_{t+l} (if t is the prediction horizon desired). We do not know the exact values of x_t, x_{t+l}, x_{t+l-1}, so we use the ones estimated by the network, given the known past values.

The model we use is a simple multilayer perceptron with one hidden layer and one output, trained by the stochastic version of the back-propagation algorithm. All input units are connected to all hidden units, and all hidden units are connected to the output unit. There is no connection between input and output units (Figure 3).

4. Comparison of the Results of Both Methods

4.1. PREDICTION OF THE ASCENDING PHASE OF CYCLE 22

We compared the different methods proposed in this paper on the prediction of the ascending phase of cycle 22. First, the McNish and Lincoln method used the data of entire cycles (case 1). Then, it used only the data of ascending cycles (case 2). The assumed last known monthly value is for July 1988. The prediction of the cycle starts in August 1988. With regard to the neural network method, RI_{12} from 1849 to July 1988 are used to train the model (training set). The data are normalized between [-1, 1] (this is a necessary step for neural networks to perform well). The prediction is made from August 1988 to September 1991 (39 values in the test set). The network is trained starting from initial random weights and the training is stopped using a cross validation method. The best network has 43 input units, 8 hidden and one output unit. We are not able to obtain good predictions with fewer input units, not enough information being present in the inputs. The parameter 6 is set arbitrarily to one. Hidden and output units use a nonlinear function: a hyperbolic tangent. We choose a nonlinear function on the output unit instead of a linear function after several trials.

Figure 4 shows the resulting curves, compared with the observed values (curve 1). Curve 2 is the result when the McNish and Lincoln method used the data of entire cycles (case 1). Curve 3 is when only the data of the ascending phase of cycles are used (case 2). The prediction based on the neural network method (curve 4) is found to be much better than the McNish and Lincoln methods for this period. The predicted maximum of RI_{12} is 154 and the predicted date is July 1989, both very close to the observed parameters (equal to 158 and July 1989, respectively). With the McNish and Lincoln methods, the dates of predicted maxima are February 1990 when entire cycles are used, and September 1990 in the other case. The maxima are overestimated and in both cases equal to 190.

The neural networks give no information on the reasons of success or failure. With the McNish and Lincoln method, poor results could at least be explained. The length of the cycle in case 1 and the length of the ascending phase in case 2 are parameters given by the average characteristics of the past cycles. When a cycle has an unusually short ascending phase, like cycle 22, the predicted maximum is obviously late. On the other hand, because of the statistical relationship between the rate of increase of the ascending phase and the maximum of the sunspot index (implicitly included in McNish and Lincoln method), the methods based on the former parameter to predict the second have failed for cycle 22 because of the unusually rapid growth of this cycle. Indeed, Lantos (1990) using a method based on the slope at the ascending inflexion point of the RI12 curve, finds 190 ± 28 (r.m.s.) with data from the minimum to July 1988, and Wilson (1990) find 185 ± 10 when using the annual rate of growth and 175 ± 30 when using the

maximum value of the difference in month-to-month smoothed RI, with data from the minimum to February 1989. Note that another attempt to predict the maximum sunspot index for the same cycle, using a neural network (Koons and Gorney, 1990), finds a similarly high value (194 ± 26). Koons and Gorney also use a multilayer perceptron with one hidden layer of units, but the methodology is different. They try directly to predict the maximum value of the cycle and the number of months from sunspot minimum to maximum and are not interested in predicting intermediate values for the ascending phase of the cycle. Their neural network has 33 inputs (3 month moothed sunspot numbers for the first 33 months of cycle 22, 17 hidden units, 2 outputs (one output unit for the maximum value and the other one for the number If months from minimum to maximum).

4.2. PREDICTION OF THE DECLINING PHASE OF CYCLE 22

We compared the different methods proposed in this paper of the prediction of the declining phase of cycle 22. First, the McNish and Lincoln method used the data of :ntire cycles (case 1). Then, it used only the data of descending cycles (case 2). The assumed last known monthly value is August 1991. The prediction of the end of l1e cycle starts in September 1991. With regard to the neural network method, RIl2 rom 1849 to August 1991 are used to train the model. The prediction is made from ~eptember 1991 to November 1994 (39 values in the test set). The best network we found (by a cross validation method) has 43 input units, 12 hidden and one output unit.

Figure 5 gives a comparison of the different predictions with the observed values (curve 1). The results with the entire cycles are given on curve 2. The results with McNish and Lincoln (case 2, curve 3) and with the neural network (curve 4) are in excellent agreement with observations. Cycle 22 as a whole is shorter than average due to a very short ascending phase of less than two years compared with four years. This explains why the results are better when the declining phase (which will have a duration of at least seven years) alone is taken into account.

The Average Relative Variance is the error test to compare the predictions with the test set:

$$\mathrm{ARV} = \frac{1}{\sigma^2}\frac{1}{n}\sum_{i\in P}(x_i - \hat{x}_i)^2$$

where σ^2 is the estimated variance of all data, P is the test set, n is the test set size, X_i is the i^{th} predicted value and x_i the corresponding desired value.

Table I gives a comparison of the Average Relative Variance for both ascending phase and declining phase, as obtained with the different methods.

Table I
Average relative variance of the different prediction methods

	ARV ascending phase	ARV declining phase
Network	0.0077	0.028
ML by using the entire cycles	0.352	0.370
ML by using separate phases	0.385	0.026

5. Conclusion

As already shown by MacPherson (1993), who applies to predictions of cycles 20, 21 and to beginning of cycle 22 the technique of neural networks and compares statistically the results with those of the McNish and Lincoln method, the former method could provide better predictions of the smoothed sunspot index than the second. We have attributed this advantage to the greater flexibility of the neural networks regarding the duration of the cycle or of the different phases of the cycles. In this respect, the atypical ascending phase of cycle 22 provides a good example. It is likely that the difference between the McNish and Lincoln method and the neural network technique will be systematically less important for cycles with more standard profiles, like cycle 11 illustrated in Figures 1 and 2 of the present paper.

References

Denkmayr, K: 1994, Thesis, Universitat Linz, Austria.
Fahlmann, S. and Lebiere, C.: 1990, in Touretzky (ed.), *Advances in Neural Networks Information Processing Systems,* Vol. 2, Morgan Kauffmann Publishers, p. 524.
Fessant, F.: 1995, Thesis, Université de Rennes, France.
Gorney, D. J., Koons, H. C., and Waltersheid, R. L.: 1993, in J. Hruska, M. A. Shea, D. F. Smart, and G. Heckman (eds.), *Solar-Terrestrial Prediction Workshop Proceedings,* Ottawa, 1992, Vol. 2, NOAA, Boulder, p. 550.
Greer, M. S.: 1993, in J. Hruska, M. A. Shea, D. F. Smart, and G. Heckman (eds.), *Solar-Terrestrial Prediction Workshop Proceedings,* Ottawa, 1992, Vol. 2, NOAA, Boulder, p. 172.
Hertz, J., Krogh, A., and Palmer, R. G.: 1991, *Introduction to the Theory of Neural Computation,* Santa Fe Institute Studies in the Sciences of Complexity, Addison Wesley, Reading, Mass.
Hildner, E. and Greer, M. S.: 1990, in R. J. Thompson, D. 1. Cole, P. J. Wilkinson, M. A. Shea, D. Smart, G. Heckman (eds.), *Solar-Terrestrial Prediction Workshop Proceedings,* Leura, 1989, Vol. I, NOAA, Boulder, p. 689.
Holland, R. L. and Vaughan, W. W.: 1984, *J. Geophys. Res.* 89, II.
Koons, H. C. and Gorney, D. J.: 1990, *EOS71,* 677.
Lantos, P.: 199O, in R. J.Thompson, D. J. Cole, P. J. Wilkinson, M. A. Shea, D. Smart, and G. Heckman (eds.), *Solar- Terrestrial Prediction Workshop Proceedings,* Leura,] 989, NOAA, Boulder, p. 492.
Lanzerotti, L. J.: 1983, *Space Sci. Rev.* 34, 347.
LeCun *et al.:* 1990, in Touretzky (ed.), *Advances in Neural Networks information Processing Systems,* Vol. 2, Morgan Kaufffmann Publishers, p. 598.
Lunstedt, H.: 1993, in J. Hruska, M. A. Shea, D. F. Smart, and G. Heckman (eds.), *Solar-Terrestrial Prediction Workshop Proceedings,* Ottawa, 1992, Vol. 2, NOAA, Boulder, p. 607.
MacPherson, K: 1993, in J. A. Jocelyn, H. Lundstedt, and J. Trolinger (eds.), 1993, *Proceedings of International Workshop on Artificial Intelligence Applications in Solar- Terrestrial Physics,* Lund, 1993, NOAA, Boulder, p. 65.
McNish, A. G. and Lincoln, J. Y.: 1949, *Transactions American Geophysical Union* 30, 673.
Ohl, A. I.: 1966, *Soln. Dann.* No. 12, 84.
Ohl, A. I.: 1976, *Soln. Dann.* No.9, 73.
Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: 1986, in D. E. Rumelhart and J. L. McClelland (eds.), *Parallel Distributed Processing,* Vol. I, MIT Press.

Stewart, F. G. and Ostrow, S. M.: 197O, *Journal des telecommunications* 37, 228.
Takens, F.: 1981, in D. A. Rand and L. S. Young (eds.), *Dynamical Systems and Turbulence*, Vol. 898, *Lecture Notes in Mathematics*, Springer-Verlag, Berlin, p. 366.
Tian, J.: 1996, G. Heckman. K Marubashi, M. A. Shea, D. F. Smart, and R. Thompson (eds.), *Solar-Terrestrial Prediction Workshop Proceedings*, Hitachi, 1996, NOM, Boulder.
Vapnik, Y. N.: 1982, *Estimation of Dependencies Based on Empirical Data*, Springer-Verlag, New York.
Wilson, R. M.: 1990, *Solar Phys.* 127, 199.
Wilson, P. R.: 1994, *Solar and Stellar Activity Cycles*, Cambridge University Press, Cambridge.